A091406528
NCC/IBL AANVRAAGBON

Verzamelnota volgt.
KOPIE PERIODIEK

30-10-2009

Datum indienen : 29-10-2009 12:17
Datum plaatsen : 29-10-2009 12:17
Aanvrager   : 0003/0133
Aanvraagident   :
Aanvragerident : 0003
Eindgebruiker   :

35556-1 IRC Int. Water and Sanitation Centre
        Library
        Postbus 2869
        2601 CW Delft
        NE
        tav

17

PPN Titel   : 220291519,1
Titel   : The American journal of evaluation
                :
Auteur   :
Deel/Supplem.   :
Corporatie   : American Evaluation Associationse:
Jaar/Editie   : 200X   Extern nummer   :
Uitgave   : [S.1.]   Elsevier Science
Serie/Sectie   :
Pag-ISSN/ISBN   :   1557-0878

Plaatscode   : 220291519 ; QA ; ; 1998 V19 - 2004 V25

Jaar   : 2002-00-00
Volume   : 23
Aflevering   : 1
Auteur   : Brown R.E., Reed C.S.ragerident. : DC 20091027
Artikel   : An integral approach to evaluating outcome evaluation
Bladzijden   : 1-17
Bron   :
Opmerking   :

An Integral Approach to Evaluating Outcome Evaluation Training

ROBERT E. BROWN AND CELESTE STURDEVANT REED

ABSTRACT

Public and private nonprofit organizations are increasing training efforts to build individual and organizational capacity to carry out and utilize outcome evaluation. Evaluators of training are challenged to find comprehensive evaluative frameworks. Traditional training evaluation tend to focus individual change, while organization-focused efforts tend to incorporate individual change as a necessary sub-component of the larger entity's change. Neither approach adequately incorporates a developmental context within the evaluative framework. This article presents an integral, developmental approach that links individual and collective attributes. The use of the framework is illustrated with examples from Check Points, an outcome evaluation training program of Michigan State University and United Way of Michigan. The article concludes with suggestions for improving training and evaluative efforts.

INTRODUCTION

Evaluative evidence of program outcomes—the human benefits that result from intervention strategies—is increasingly demanded from public and private nonprofit organizations (Horsch, 1996). In addition to utilizing external evaluators, organizations are responding to

Robert E. Brown • University Outreach Partnerships, Michigan State University, 6 Kellogg Center, East Lansing, MI 48824-1022, USA; Tel: (1) 517-432-2500; E-mail: brownr23@msu.edu.

accountability pressures by building their own capacity to generate evaluative evidence of client outcomes (Plantz, Greenway, & Hendricks, 1997). To carry out and utilize outcome evaluation, they are turning to training as a strategy to build evaluative competency and capacity both within individual employees and across the organization. Evaluators of training, in turn, are challenged to find comprehensive evaluative frameworks that embrace this holistic context.

For the last 40 years, the pre-eminent framework for evaluating training effectiveness has been the model developed by Donald Kirkpatrick (Alliger, Tannenbaum, Bennet, Traver, & Shotland, 1997; Bumpass, 1990; Clement, 1982). In this sequential model, training programs are evaluated on one or more of four criteria—reaction, learning, behavior (application or transfer), and results (Kirkpatrick, 1959, 1998). While this approach (Alliger et al., 1997) can serve as one framework to examine training effectiveness, it has a number of significant limitations. For example, it provides little illumination of other variables (beyond training) that could also influence the transfer of learned skills to on-the-job settings and subsequent organizational results. Many variables can affect skill transfer and organizational results including: Trainee readiness for the training; trainee motivation; opportunities for practice and feedback during the course; lack of similarity between the training setting and the job setting; lack of opportunities to apply the training on-the-job; the internal organizational environment, especially boss, peers, organizational policies; and the external environment (Baldwin & Ford, 1988; Clement, 1982).

Approaches for altering an organization's performance (i.e., results) or character (the internal organizational environment, including opportunity to transfer skills) have generally been within the domain of organization development or organizational change interventions, rather than training. Organization development models promote change through four primary interventions, categorized as: process, technostructural, human resource or strategic interventions (Cummings & Worley, 1997). These models acknowledge the importance of focusing on multiple levels—individual, group, and/or organization—although the interventions may emphasize one more than another.

Organization-focused efforts tend to incorporate individual change as a necessary subcomponent of the larger entity's change; in contrast, traditional training initiatives tend to view organization improvements as a by-product of individual change. Increasingly practitioners and researchers are focusing their efforts on understanding and evaluating the processes by which individuals *and* organizations strengthen their abilities to carry out their functions and achieve desired results—the phenomena of capacity building or capacity development (Morgan, 1997). Capacity building research emphases many different aspects of this phenomenon. For example, investigations of social capital highlight the individual networks, norms and trust that facilitate cooperation (Onyx & Bullen, 2000). Organizational learning focuses on the widespread collection and use of internally generated information to transform organizations (Senge, 1990). Whatever the particular emphasis, capacity building connotes simultaneous development on multiple levels.

In this article, we present an integral approach as one such comprehensive evaluative framework that embraces holistic context. As an illustrative case, we will turn to Check Points— a training program that responded to the desire to build outcome evaluative competency and capacity both within individual employees and across the organizations.

AN INTEGRAL FRAMEWORK

One holistic approach used to link individual and collective attributes (MSU Asset Work-group, 2000) is an integral framework (Wilber, 2000). There are four key concepts to this framework:

(1) development is nested;
(2) it occurs in four distinct but inter-related realms;
(3) within each realm there are multiple developmental pathways; and
(4) between realms there are interactions.

The first concept, nested development, is the notion that each level of development is whole unto itself and part of something else. Here, it is applied to the person–organization relationship. Thus, a training participant is whole unto herself and also part of a work group, which is whole unto itself and part of a department, which is whole unto itself and part of an agency.

The second key concept is that development must be understood from four distinct but inter-related realms (Wilber, 2000). These are the realms of intention, behavior, culture, and so-cial structure which, taken together, encompass the interior (subjective) and exterior (objective) space of both the individual and the collective. In this framework, development is understood by simultaneously considering the interaction among these four aspects of the smaller and larger units:

Realm of intention—the beliefs, understandings and values that individuals (work groups, departments; the smaller unit) must hold.
Realm of behavior—the skills and the practices that individuals (work groups, departments, etc.) must adopt.
Realm of culture—the mutual understanding and agreement that must exist among members of the larger unit (the work group, the department, the organization).
Realm of social structure—the structure that must exist within the larger unit (workgroups, departments, and the organization as a whole).

The third key concept inherent in the framework is that *within* each realm there are multiple developmental pathways. Research has confirmed some of these pathways—cognition, morals, creativity, affects (Wilber, 1999). In other instances, practice has provided insight into these pathways, as seen in the suggestions for adopting outcome evaluation (United Way of America, 1996). Without regard to the method of verification—research or experience—each step must be developmental; i.e., it transcends and includes the prior step.

The final integral framework concept links developmental progress in one realm to de-velopmental progress in another. Again, research can illuminate relationships among the four realms. For example, based on prior accumulated research, the following links are now gen-erally accepted. A person who thinks a new skill will be useful is more inclined to acquire it (Cheng & Ho, 1998; Noe & Schmitt, 1986). A person is more likely to transfer those new skills to their job when their supervisor expects they will use that skill, and especially when the use of the skill will be monitored in their annual performance review (Facteau, Dobbins, Russell, Ladd, & Kudisch, 1995; Rouiller & Goldstein, 1993). Thus, individual understanding is linked to behavior; individual behavior is linked to mutual understanding, and mutual understanding is reinforced by organization structures. Although this example suggests a sequenced order among the realms, that is not necessarily the case. The framework provides a constant reminder

of likely simultaneous occurrence across all the realms; research and/or experience can be used to verify the relationships among the realms.

AN INTEGRAL FRAMEWORK FOR BUILDING OUTCOME EVALUATION CAPACITY

An integral framework provides the basis for conceptualizing developmental pathways that operationalize outcome evaluation within organizations. To start, nested levels or stages, upon which development is built, must be discerned. The work of Plantz et al. (1997) offers such a view for outcome evaluation. They posit that outcome evaluation capacity moves through the stages of getting ready, choosing outcomes, and choosing indicators, through collecting and analyzing data, to using findings for program improvement.

Table 1 represents Plantz's stages elaborated across and within the four realms of an integral framework.

It articulates key aspects present within each realm at a particular level of outcome evaluation capacity (Andrews, Reed, Brown, Hembroff, Seabrook, & Villarruel, undated; United Way of America, 1996). When read within a realm, it provides a sequential map of expanding comprehensiveness. When read across a stage, it provides corresponding points of relationship among realms. For training programs aimed at building evaluative competency and capacity both within individual employees and across the organization, consideration of the within realms plus across stages is important.

The increasing comprehensiveness outlined in the intentional and behavioral realms of Table 1 reveal both knowledge and skill competencies that individuals need to do outcome evaluation and also sequential patterns for organizing training content. The increasing comprehensiveness in the cultural and social realms reveal a road map for operationalizing outcome evaluation across organizations as well as clarifying additional training topics beyond individual competency and skill. The corresponding points of relationship between realms show aspects that must be actualized before outcome evaluation capacity may be realized within individuals and across the organization at particular levels. If training is a change strategy to build this dual capacity, it must take into account both the within and across view. Likewise, evaluations of training effectiveness must do the same.

USING AN INTEGRAL FRAMEWORK TO EVALUATE OUTCOME EVALUATION TRAINING: AN ILLUSTRATIVE CASE STUDY

The use of the integral framework for evaluation will be illustrated with data and anecdotes collected from several outcome evaluation training series. In 1997, Michigan State University Outreach Partnerships and United Way of Michigan collaborated to create Check Points, a training program designed to increase the capacity of health and human services professionals to do outcome evaluation. A university–community team was assembled to provide guidance on the selection of course content and a small group of MSU Faculty and United Way staff designed and provided the training. Many evaluation topics, some familiar and some new, were introduced during the training series. Those agency representatives who participated in the design of the series assumed that participants could be introduced to focal concepts and immediately put those concepts to use in their agencies. In order to increase

TABLE 1.
A Developmental Path within Realms to Operationalize Outcome Evaluation

Stages	Intention	Behavior	Culture	Social Structure
1. Haven't started	No individual belief on the importance of outcome evaluation Lack of evaluation skills	Lack of evaluation practice or behavior	No mutual understanding of outcome evaluation No mutual agreement on its importance	Lack of evaluation structures and tools
2. Get ready	Understand key terms Personal commitment Time commitment	Assemble and orient outcome work team Adopt timelines	Mutually understand and agree upon expectations and plans Peer, management, organizational commitment Resource/time commitment	Timelines for installing outcome evaluation Work sessions
3. Choose outcomes	Understand relationship between activities and initial, intermediate, long-term outcomes	Construct logic models Adopt logic models	Mutually agree on outcomes to measure	Logic models Work sessions
4. Specify indicators for outcomes	Understand what constitutes an indicator	Specify one or more indicators for each outcome	Mutually agree upon indicators specified	Evaluation plan Work sessions
5. Prepare to collect data	Understand data sources Understand data collection methods and instruments	Identify data source for chosen indicators Design data collection methods and tools	Mutually agree on data sources, collection methods, tools	Evaluation plan Data collection tools Data collection procedures Data storage

TABLE 1.
Continued

Stages	Intention	Behavior	Culture	Social Structure
6. Try out outcome measurement system	Understand the importance of a trial run	Conduct a trial run	Mutual agreement on resource allocation for a trial run	Evaluation plan Data collection tools Data collection procedures Data storage B statistics program
7. Analyze data and report findings	Understand data analysis strategies and techniques Understand reporting methods and formats	Analyze data Report findings	Mutual agreement on type of data analysis, report items and format Mutual agreement on resource allocation to do analysis and publish report	Data storage B statistics program Report software
8. Improve measurement system	Believe that continuous improvement is important Understand results of the trial run	Enact improvement strategies Continue outcome evaluation efforts	Mutually agree on what the trial run tells us and subsequent improvement strategies Mutual agreement on resource allocation for continued outcome evaluation	Continuous improvement documents and structures Evaluation plan Data collection tools Data collection procedures Data storage B statistics program
9. Use findings	Understand relationship between findings and program interventions	Determine and enact intervention improvement findings	Mutually agree on what the findings tells us and subsequent intervention improvement strategies	Program management structure Program intervention structure

Behavioral steps adapted from United Way of America.

the back-on-the-job usefulness of the content, participants were to develop an individualized evaluation plan.

The series was not associated with a degree granting program or other formal curriculum, nor was it a required component of any organization's professional development program. From the University's standpoint, the participants were voluntary "explorers" who could take information back to their agency and apply it. From the United Way of Michigan's standpoint, their practitioner participants were potential emissaries to local organizations, many of which needed to be convinced that an outcome evaluation approach was a good thing to do. Both partners recognized that participants' beliefs in their own knowledge and skill acquisition as well as the content's potential applicability would be essential for accomplishing these compatible but separate ends. The emphasis on these participants' understandings also shaped the session reaction forms.

The training series were delivered over the course of 3 years. As the designers and principle trainers of Check Points, the authors were interested in the extent to which participants felt they had gained information and skills in the sessions and also expected to transfer their knowledge and skills to their jobs. Organizational and contextual data were collected as potential explanatory factors for trainees' back-on-the-job use. Over time, this training intervention has been more fully conceptualized as a developmental change process to create capacity-building by using an integral framework. That reconceptualization is reported here.

METHOD

Findings reported here are based on data collected from participants who attended one of the five different series of Check Points Training. The format of these training series varied: two cohorts participated in 1 day/month series; one cohort participated in week-long series; and two cohorts participated in alternate formats. Participants held executive, administrative, supervisory or direct service positions in a variety of educational, health and human services agencies, such as University Extension, Local Community Public Health Department, and Boys and Girls Clubs. At an initial training session, participants completed a "Background Information" instrument designed to collect data on variables such as their prior skills and interests, the culture of their organization, and the pressure from other organizations to move toward an outcome orientation. At the end of each training day, participants completed reaction forms that asked them to assess their gain in knowledge and the potential usefulness of the content delivered. Follow-up surveys were sent to participants in the first two cohorts. All data were collected using self-report instruments rather than knowledge tests (recall the perception of gain and usefulness discussed above). When self-report instruments are used to assess pre-training knowledge, participants are likely to overestimate the amount they know. More realistic assessments at the end of a session may therefore, show "no gain" or even a loss in knowledge. To compensate for this phenomenon, we used a form of pre/post-test on the end of session reaction form (Pratt, McGuigan, & Katzev, 2000; Rockwell & Kohn, 1989). The first of the paired items asked people to think back *prior to the session* and rate themselves on the itemized content presented. The second paired item asked them to rate themselves *now* (at the end of the session).

Thirty-five content items were rated by participants. All paired items were rated using 5-point Likert scales with 5 = very/most and 1 = not at all. Gains were determined by comparing the pre- and post-test means. The level for determining statistical significance

was set at $p > .05$. For this effort, identification of some level of practical significance, not just statistical significance, was desired. In some venues there has been sufficient research to set standards for practicality; in most cases such standards are nonexistent. Therefore, we arbitrarily set a level for practical significance equal to a change in the mean that was at least one standard deviation higher from the pre- to the post-test. This seemed "big enough" to capture meaningful change in participants' self-reported understanding and knowledge.

FINDINGS

Our findings are presented by realm. Because the evaluation of the Check Points training program serves to illustrate the use of an integral framework, we have provided summary findings rather than traditional results based on data. Readers interested in numerical results are encouraged to contact the authors. The training content emphasized four of the specific stages presented in Table 1. Those stages are reproduced in Table 2; the associated training session is identified as are training session practice items and products.

We begin with the two realms that address individual change: the intentional and the behavioral realms.

Intentional Realm

Within this realm, partners were interested in two questions: Did participants believe that increased understanding occurred? Did Check Points participants see utility—a belief in transferable value—in the training?

Understanding. Significant gains were found for 34 of the 35 content items; only "identifying the links in our program" did not show significant improvement. Although there was a large difference between the pre- and post-score means, the huge standard deviation suggested that answers ran the entire gamut from *none* to a *lot*. Our experience supported this interpretation. In this content area, participants constructed maps or mental models of their programs and were then asked to discern distinctive paths or connective links within these pictures. For most participants this was a novel process that created both cognitive and application struggles.

While it was rewarding to find that the participants reported gaining knowledge overall, we were especially interested in those areas in which the gains may have been large enough to suggest practical improvements. Twelve items, about one-third (35.3%) of the content, had substantial gains. Those items clustered in sessions as follows:

Session #2: one of six content items, identifying the logic in our program;
Session #3: two of six content items, both related to identifying appropriate methods;
Session #4: five of nine content items, all related to improving surveys and their use;
Session #5: two of seven content items, data processing and research design;
Session #6: two of three content items, cost estimation of evaluation and cultural competence.

To more fully understand the nature of the reported gains, content items were also ranked in ascending order. Little change was seen in several of the sessions. Session 1 (an overview of evaluation) content items started high and ended high; most participants were familiar with

TABLE 2.
Sessions with Training Program Practice Items and Products by Stage and Realm

Stage	Intention	Behavior	Culture	Social Structure
3. Choose outcomes [Session #2]	Understand relationship between activities and initial, intermediate and long-term outcomes	**Construct logic models**	Mutual agreement on outcomes to measure	*Logic model*
4. Specify indicators for outcomes [Session #2]	Understand what constitutes an indicator	**Specify one or more indicators for each outcome**	Mutual agreement on indicators specified	*Evaluation plan*
5. Prepare to collect date [Sessions #3–5]	Understand data sources	**Identify data source for chosen indicators**	Mutual agreement on data sources, collection methods and tools	*Evaluation plan* (continued)
	Understand data collection methods and instruments	Decide on data collection methods and design tools		Data collection tools
				Data collection procedures
				Database
7. Analyze data and report findings [Sessions #5 and 6]	Understand data analysis strategies and techniques	Analyze data	Mutual agreement on type of data analyses, report items and format	Database and statistic (or other analysis) program
	Understand reporting methods and formats	**Report findings**	Mutual agreement on resource allocation to do analyses and public report	Report software

Note. Session #1 was a persuasion session, focusing on the need for outcome evaluation and the attributes of a learning organization. Items in bold were practiced in the training sessions; drafts of the items in italics were produced in the training.

the ideas presented in this session. In Session 5, which dealt with data analysis and evaluation design, the content items ranked at the low end of both the pre- and post-ranked order. Given the prior experiences of these direct service providers and managers, this is not a surprising finding. Of all the training sessions, participants voiced the greatest apprehension about these topics. Sessions 3 and 6 also showed only marginal movement. Session 3 content items that dealt with methods and measures remained heavily distributed in the top two quartiles. Session 6, a wrap-up session that dealt with costs, reporting and other unfinished issues—maintained a wide spread from bottom to top.

Session 4, focused on data collection methods, showed much movement. Seventy-five percent of the before-session content areas were ranked in the bottom half compared to 62% of the after-session content which was ranked in the top half. In addition, almost half (5) of the 12 content items that were found to have increased in practical significance were located in Session 4.

In summary, the statistical analysis of the data produced three major findings. First, participants' improved understandings occurred across the entire content of the training program. This program contained substantial novel information. Second, those 12 items that achieved practical significance addressed three of the important understandings on our hypothesized developmental path (cf., Table 1 or 2). People reported substantial gains in understanding the logic of their programs, appropriate data collection methods and appropriate data collection tools. Third, while gains were reported by participants, there was substantial room within the content for additional understanding to occur.

Utility. To understand utility we asked: How applicable to your situation was today's session? This was an important question because all partners expected participants to use the information. Participants rated four responses: the content

- is directly applicable to my current work;
- is more applicable to others in my organization than it is to me;
- is not currently applicable, but fits with the direction that my agency is moving toward;
- is useful to me in other things I am doing outside my agency.

The responses indicated that a substantial majority (80% or more) of participants felt that the training content related to their own work. Only in Session 5 was there a lack of a clear majority. Session 5, as stated previously, dealt with data processing and analysis, along with evaluation design. The finding is congruent with the positions held by most of the participants. Rarely was there a participant whose job required them to do data analysis; most had direct service, program supervision or broad management positions. As further verification, approximately 30% felt that the content in that session was more applicable to others in the organization.

Responses to the third measure of utility—"is not currently applicable, but fits with the direction that the agency is moving toward"—gave no conclusive meaning. Only in Session #2, where logic models were introduced, did a substantial number (60%) of participants give a positive response. A negative response could indicate that *either* the agency has enacted outcome evaluation using that particular content *or* the content was not currently applicable and the agency was not moving toward it. An examination of follow-up data, discussed below, provided insight into these answers, suggesting that people probably felt the content was applicable, but they had not progressed to the stage where it would be applied.

Behavioral Realm

There are both anecdotal and empirical data that provide insight into the acquisition and transfer of outcome evaluation skills. First, we will report our impressions of participants' experiences applying the content in the training sessions; this is followed by the results of our follow-up surveys.

Opportunity to practice using the skills as well as applying the content to participants' specific situations was built in throughout the training program. In particular, a substantial amount of time was spent constructing and refining logic models, first in order to convey the critical elements of an intervention and then to serve as a guide for developing an evaluation plan. For successful practice, participants had to grasp the essentials of a new outcome language and a new way of conceptualizing their programs. Participants differed in their willingness to try out the content on their own programs, their ability to grasp the concepts and their facility in translating the concepts into logic models. People did not need to turn in their assignments at the end of the training day, but it was our impression that most of the teams left with rudimentary models that could guide at least one of their members' work.

Transfer of evaluation skills was assessed with a follow-up survey. At the beginning of the first session, participants were asked whether they or someone else in their organization were engaged in any of 17 basic evaluative roles. Six months after the completion of Check Points, a subset of the United Way-related participants were again asked to rank the status of these roles. Participants were asked to check one of the following categories: you perform, someone else performs, you both perform, neither performs. Increases in either or both individual and organization capacity can be inferred from these responses:

- the ranking of *neither performs* should decrease over time;
- *you perform* should increase as people transfer new skills to their jobs;
- *someone else performs* should increase as the practice of doing outcome evaluation activities spreads through the organization; and
- *you both perform* should increase if individual and organizational capacity building are occurring.

Different evaluation roles were being carried out at different rates. Conducting a needs assessment was an evaluation role that was being carried out in almost all (82%) of the participants' organizations. Initially, the evaluation roles least likely to be performed by anyone were:

- revise an evaluation plan based on how it worked (43%);
- design an evaluation for my program (41%);
- work with an evaluation consultant as a liaison (40%);
- develop an evaluation report (38%);
- design an evaluation plan for the agency (35%);
- select measures that fit our clients and programs (34%);
- analyze case records (33%);
- use evaluation results to improve my program (32%).

Six months after training, the *neither performs* responses for 15 out of 17 roles had decreased from the pre to the post; *you perform* responses increased for 9 out of 17 roles; *someone else performs* responses increased in 13 out of 17 roles; and *you both perform* responses

increased in 17 out of 17 roles. These findings suggest individual and organization capacity
has increased.

For confirmation of suggested gains in organizational capacity, we turn to the remaining
two realms—cultural and social structural—that address organizational change.

Cultural Realm

The cultural realm encompasses the attributes of mutual understanding and agreement
within groups, organizations, communities, etc. Mutual understanding and agreement must be
present for organizations to progress along a developmental path from supporting outcome
evaluation to adopting the practices and using the results.

At the 6 month follow-up, participants were also asked to check the stages they had
completed in implementing outcome evaluation in their organization (refer back to Table 1).
These stages represent a developmental view of organizational capacity to utilize outcome
evaluation. Thirty-one participants responded:

- 39% perceived themselves at various stages of developing their evaluative systems
  (stages 2–5);
- 30% perceived themselves at the stages of trying out and improving their evaluative
  systems (stages 6–8); and
- 32% perceived they were using evaluative findings to improve interventions
  (stage 9).

The same sub-set of participants were asked to indicate which, if any, of eight factors
might have influenced their implementation efforts:

- I haven't found the time;
- I'm not comfortable with my ability to do it;
- it wasn't practical for my situation;
- I tried it and it didn't work;
- peers aren't motivated to try it;
- subordinates aren't motivated to try it;
- there was limited support from my bosses;
- it is not an organizational priority.

This list includes both individual factors (wasn't comfortable) and organizational factors
(no time, not practical, didn't work, peers/subordinates weren't motivated, limited support, not
an organizational priority).

The following picture emerges from our analyses.

- Almost all of these respondents have started through the organizational stages to install
  outcome evaluation approaches in their own organizations. Even those who say they
  haven't started are at the stage of *getting ready*.
- Not everyone who had started was moving forward. A non-supportive organizational
  culture was reported by some people. In this culture, people reported that while they
  had started, they found that it wasn't practical and their actions had limited support
  from their boss. That lack of practicality was related to a previously reported lack
  of organizational priority. Finally, the absence of an organizational priority was not
  only related to limited support from one's boss but also to their peer groups' lack

of motivation. Where peers weren't motivated to use the information, neither were subordinates.

- A supportive culture was reported by other people. On the background survey, participants previously reported the following: an organizationally endorsed shift toward focusing on results; their organization expected the participant to share the training information with others; and the organization was open to learning from evaluation results. This willingness to learn carried forward in time to the reported "support from one's boss". Finally, support from one's boss was associated with "having started".
- "Getting ready" was related to a prior organizational shift to focusing on results. That shift may have been the result of primary funders requiring a focus on outcome evaluation. Where funders required outcome evaluation, so did the organization. It was interesting to note that this alignment of organizational and external forces was not associated with any more advanced stages.
- Some people didn't progress beyond trying to "choose outcomes" because they "tried and it didn't work".
- If people got beyond trying to choose outcomes, the stages were related to each other and seem to be increasingly integrated. Where people got as far as specifying indicators, they were also preparing to collect data. When people got as far as trying out the measurement system, they were also trying to improve the system and they were using the findings from their measurement system to improve their interventions.
- The only two influencing factors that were not related to some stage of outcome evaluation implementation were the participants' lack of time and their lack of comfort with the content.

Social Structural Realm

No structural-related data were collected. This was not a planned omission, but rather reflects the absence of this integral framework at the time of data collection. We do know that some participants left the training program with draft logic models and evaluation plans that could serve as discussion documents. No data were collected about organizations' structural changes that may have supported the adoption of outcome evaluation methods.

SUMMARY

Linking the Realms of Intention and Behavior

The first insight has to do with the link between gains in individual's knowledge understanding and their ability to translate that new knowledge into practice. People did report gains from their experiences. However, there were mixed results from their 6-month reports. Recall the relationships between "have started" and "wasn't practical" and also between "choosing outcomes" and "tried but didn't work". When viewed together, these support what trainers in the program have come to understand over time, namely, that there is a double learning curve (Fig. 1).

Acquiring outcome evaluation proficiency is a complex process. Many participants were learning a new way of thinking about their programs and a new language for outcome evaluation. At first participants had a difficult time framing and planning an outcome evaluation
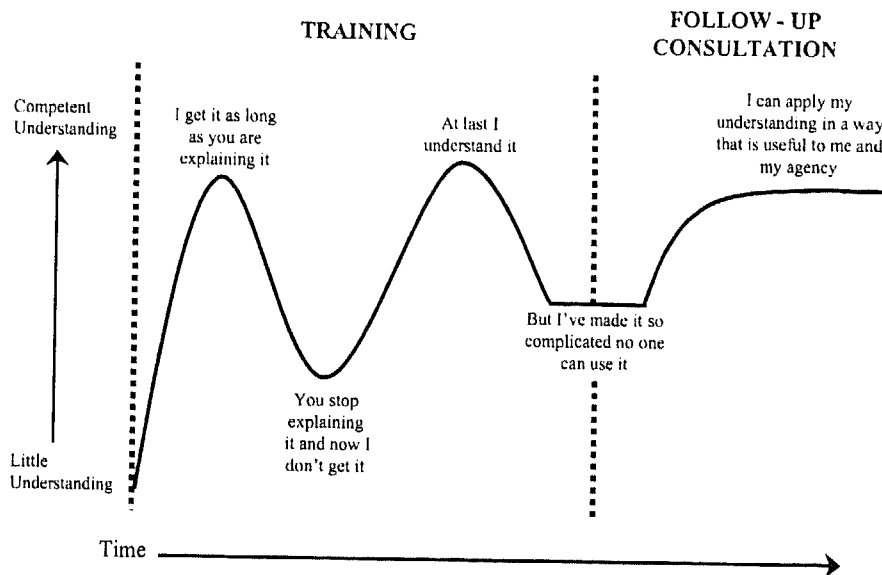
Figure 1. The Learning Curve of Outcome Evaluation.

without the direct assistance of the trainer. During the sessions, as knowledge and comfort increased, participants were able to embrace these activities with less and less assistance—the apex of the first learning curve. In most instances, the resulting frameworks and plans were so complicated that it would be impractical to use on-the-job. Being able to simplify one's design and implementation plan represents the knowledge and skills of the second learning curve. We interpreted the "tried and didn't work" or "impractical" as one sign respondents had not moved beyond their too complicated plans.

Linking the Realms of Behavior and Culture

Most of the participants were sponsored by their employer and there seemed to be general agreement among them that they would be applying the information gained on-the-job. One of the requisites of successful application of outcome evaluation practice is agreement that evaluation is an important—even necessary—component of organization action. The decision to begin evaluating programs is rarely based on a single individual's decision. Several of the participants were themselves agency directors, but even they would need staff and/or board agreement in order to initiate the adoption of evaluative practices. The culture relationships noted above, e.g., between "have started" and "not an organizational priority", suggests that for some of these participants there was not common organizational agreement about what using these skills would mean for the agency.

Linking the Realms of Behavior and Structure

Almost one-third of the participants (32%) reported that they had progressed to using the findings from their evaluative practice to improve agency practice. This response connotes

a feedback process, or structure, for incorporating data. While some of the participants may have developed evaluation plans for their own work, most in the sessions were engaged in developing plans that affected projects or programs—other people's work. That suggests that there were some supportive social structures.

DISCUSSION

In this article, we have illustrated how an integral framework can be used to judge the comprehensiveness of a capacity-building strategy aimed at achieving individual and organizational gains. In concluding, we offer four ways in which the framework can also be used to improve capacity-building efforts.

First, an integral framework helps us understand the unfolding nature of developmental stages within and across realms. That unfolding, in turn, presents a range of possible points for evaluation. This is important for two reasons. Knowledge of appropriate developmental pathways provides an alternative to "all or nothing" indicators for achieving an outcome. Improved academic achievement for a child being tutored, e.g., may be appropriately measured with "homework regularly completed" or "increased number of right answers on tests" as evidence of intermediary accomplishments on the path to "improved grade in XX class". This is especially important when interventions lack sufficient dose or duration to reasonably achieve desired outcomes (cf., Scott & Sechrest, 1989). Recognition that development does not occur uniformly across realms prepares partners for streams within different realms to progress at different rates (Wilber, 1999). An integral framework may serve as a useful tool with training stakeholders to identify key evaluative points along particular paths that would suggest movement toward desired individual and organizational capacity gains. It also helps us clarify the types of evaluative evidence that might be persuasive to various stakeholders. For example, in our outcome evaluation training, the United Way partners saw the agency participants' *perception* of knowledge gained and utility as meaningful indicators of training success. The partners recognized that an initial task for participants was likely to be persuading others in their organizations that outcome evaluation was an important effort. Without gaining this mutual understanding and agreement in the organization, participants were unlikely to be able to apply their new knowledge and skills.

Second, using an integral framework provides a basis for conversations with training stakeholders concerning potential time frames needed to gain competencies/capacity across all realms for a specific stage. For example, an individual may gain basic skills in logic modeling faster than agreement can be reached to use those skills in their work unit or organization. Where stakeholders are interested in evaluating training application success, they must first understand that application depends on not only on individuals' skill attainment but also organizational agreements to apply those skills and the social structures to support the use of those skills (cf. the literature on transfer of training). They must also come to some understanding of the length of time, beyond the formal training periods, it might take within the organization to achieve mutual understanding and agreement that will allow for application. These types of discussions should aid evaluators with the timing of follow-up data. Recall that at 6 months, those people who were preparing to collect data (Stage 5) were doing so within an organization that "expected them to share what they had learned". Others who lacked that organizational expectation had not progressed to that stage.

Third, an integral framework helps us judge the comprehensiveness of our training content. The developmental stages, across the realms, of any training goal provides a road map for training content. Individual competency and capacity centers on the realms of intention and behavior. We designed this outcome evaluation training using university content experts, relying on the best practices in the field developed by the United Way of America, and building practice time into our training sessions so that participants would have an opportunity to apply the content. In doing so, we concentrated on the understandings, knowledge and skills required to *use* outcome evaluation on-the-job, the intention and behavior realms pathways. There was even some discussion of the social structures (i.e., data systems) that would need to be available to support outcome evaluation efforts. However, we paid scant attention to the types of persuasive arguments that participants might need to convince their colleagues. We depended on the untested assumption that since their organizations paid for the training, agency conditions were supportive. Using the integral framework, we might have augmented the training content with more attention to companion cultural stages, perhaps eliciting and/or providing scripts that participants might use with their colleagues. As external trainers, we had little opportunity to directly address organizational capacity elements. However, the framework could have been used to target and improve our follow-up investigation of culture and social structure elements.

Finally, an integral framework helps trainers think about the strategic mix of participants given the stages of development that training is addressing. Those of us providing external training rely on messages that specifically identify appropriate audiences. Where the intent is to change the practice of a group or organization, the recruitment messages say "a team representing . . . is most appropriate"—and group training rates are offered. In the outcome training program reported here, we relied on our community partners to identify (or self-select) appropriate participants from whom we collected background and follow-up information on the evaluative roles they played. We were trying to select people who would be responsible for enacting outcome evaluation, not those who would smooth the way for such practices to occur. Some agency directors participated because they were interested in the content. During the design phase of this training, a brief overview of the content was considered for board members and executives; think of the proposed sessions as the equivalent to "Accounting for Non-Financial Managers". The decision was made to start with professional staff who would be expected to implement outcome evaluation practices. That approach achieved certain results, reported here, but attention to the integral framework would suggest that we were on the right path in considering the overview sessions.

The use of an integral framework, illustrated here with an outcome evaluation example, can be used not only to guide evaluation but also to improve training design and delivery. It can serve to focus conversations for audiences concerned with monitoring individual as well as organizational change. And, in particular, the framework can assist in understanding the complexity of not only modest but comprehensive change strategies.

REFERENCES

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relationships among training criteria. *Personnel Psychology, 50,* 341–358.

Andrews, M. P., Reed, C. S., Brown, R. E., Hembroff, L. A. Seabrook, L., & Villarruel, F. A. (Undated). Check Points—a training manual for the evaluation of educational and human services organizations. East Lansing, MI: ADS@MSU and United Way of Michigan.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: a review and directions for future research. *Personnel Psychology, 41*, 63–105.

Bumpass, S. (1990). Measuring participant performance—an alternative. *Australian Journal of Educational Technology, 6*(2), 99–107.

Cheng, E. W. L., & Ho, D. C. K. (1998). Transfer of training: some practical thoughts from theoretical studies. *International Journal of Management, 15*(1), 14–19.

Clement, C. W. (1982). Testing the hierarchy theory of training evaluation: an expanded role for trainee reactions. *Public Personnel Management Journal, 11*, 176–184.

Cummings, T. G., & Worley, C. G. (1997). *Organization Development and Change* (6th ed.). Cincinnati, OH: South-Western College Publishing Company.

Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management, 21*(1), 1–25.

Horsch, K. (1996). Results-based accountability systems: opportunities and challenges [6 paragraphs]. *The evaluation exchange: emerging strategies is evaluating child and family services* [On-line serial], 2(1). Available FTP: Hostname: hugse1.havard.edu Directory: /~hfrp/eval/issue3/horsch.html.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of ASTD, 13*(11), 3–9.

Kirkpatrick, D. L. (1998). *Evaluating training programs* (2nd ed.). San Franciso: Berrett-Koehler.

MSU Asset Workgroup (2000). Michigan State University's Asset/Outcome Workgroup Model: Overview of a Work in Progress. Working Paper. East Lansing, MI.

Morgan, P. (1997). The design and use of capacity development indicators. Paper prepared for the Political and Social Policies Division, Policy Branch, CIDA. Que., Canada: Canadian International Development Agency (CIDA).

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: test of a model. *Personnel Psychology, 39*, 497–523.

Onyx, J., & Bullen, P. (2000). Measuring social capital in five communities. *Journal of Applied Behavioral Science, 36*(1), 23–42.

Plantz, M. C., Greenway, M. T., & Hendricks, M. (1997). Outcome measurement: showing results in the nonprofit sector. *Outcome Measurement Resource Network*. Available: http://www.unitedway.org/outcome/ndpaper.htm.

Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: using retrospective pretest measures. *American Journal of Evaluation, 21*(3), 341–349.

Rockwell, K., & Kohn, H. (1989). Post-then-pre evaluation: measuring behavior change more accurately. *Journal of Extension, 27*(2).

Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. *Human Resource Development Quarterly, 4*(4), 377–390.

Scott, A. G., & Sechrest, L. (1989). Strength of theory and theory of strength. *Evaluation and Program Planning, 12*, 329–336.

Senge, P. M. (1990). *The fifth discipline: the art and practice of the learning organization*. New York: Doubleday.

United Way of America (1996). *Measuring program outcomes: a practical approach*. Alexandria, VA: Author.

Wilber, K. (1999). *The collected works of Ken Wilber: integral psychology, transformations of consciousness & selected essays* (Vol. 4). Boston: Shambhala Publications.

Wilber, K. (2000). *The collected works of Ken Wilber: sex, ecology, spirituality* (Vol. 6). Boston: Shambhala Publications.